

A Statistical Analysis of the Determinants of Electricity Pricing in the United States: Regional, Sectoral, and Demographic Perspectives

Amrit Subedi¹

¹Department of Mathematics and Statistics, College of Arts and Sciences,

University of South Florida, Tampa, Florida, USA,

tirma321@gmail.com/amritsubedi@usf.edu

Jul 2026

Abstract

The United States electricity market is a complex and dynamic landscape influenced by a multitude of demographic, geographic, and infrastructural factors. This study investigates the primary determinants of electricity pricing in the U.S., utilizing cross-sectional data from the U.S. Energy Information Administration (EIA) and the U.S. Census Bureau for January 2023 and January 2024. A multiple linear regression approach, incorporating a Box-Cox transformation to satisfy normality assumptions, was employed to analyze the impact of sector, geographic region, customer volume, and customer density on electricity prices. The final log-transformed model explains over 85% of the variance in pricing (Adjusted $R^2 = 0.8536$). The findings reveal that commercial and industrial sectors benefit from significantly lower prices compared to the residential sector. Geographically, the Northeast exhibits the highest pricing, with the Midwest, South, and West regions showing significantly lower rates. Furthermore, while the absolute number of customers has a weak impact, customer density positively and significantly influences pricing, likely due to urban grid infrastructure costs. Temporal variations between 2023 and 2024 were found to be statistically insignificant. These findings offer valuable insights for policymakers and utility stakeholders aiming to optimize pricing structures and manage infrastructural density.

Keywords: Electricity pricing, energy economics, multiple linear regression, regional disparities, customer density, box-cox transformation

1. Introduction

The electricity market in the United States operates within a highly dynamic environment, governed by varying regional demographics, infrastructure constraints, market regulations, and sectoral demands. Understanding the dynamics of electricity pricing is crucial not only for industry stakeholders seeking to optimize grid operations but also for policymakers aiming to protect consumers and ensure equitable access to energy.

Despite the significance of electricity pricing, there remains a need for a holistic understanding of the concurrent factors driving price variations across different sectors and regions. While existing literature provides valuable insights into individual aspects of energy economics—such as the impact of fuel costs or deregulation—a comprehensive statistical analysis that simultaneously integrates regional geography, sector types, customer volume, and geographic density is less common.

This study aims to address this gap by investigating the collective impact of these predictor variables on average electricity prices across U.S. states. The primary objectives are: (1) to compare electricity pricing across different sectors and regions, (2) to assess temporal pricing differences between 2023 and 2024, (3) to establish the mathematical relationship between demographic predictor variables and pricing, and (4) to explore the interaction effects of these factors.

2. Literature Review

The determinants of electricity pricing have been extensively studied within energy economics, typically categorized into supply-side factors (fuel costs, generation mix, grid infrastructure) and demand-side factors (consumer behavior, density, and sectoral load profiles).

2.1 Regional Disparities in Pricing

Geographic location plays a pivotal role in electricity pricing due to regional variations in resource availability, generation constraints, and regulatory frameworks. Borenstein and Bushnell (2015) highlighted that regional grid interconnects in the U.S. experience vastly different marginal costs based on their primary generation fuels. For example, the Northeast region has historically faced higher electricity prices due to infrastructure constraints in natural gas pipelines and a heavier reliance on expensive peak-generation during extreme weather (Joskow, 2006). Conversely, regions like the Midwest and South often benefit from lower prices due to closer proximity to natural gas and coal production, as well as less stringent regulatory costs (Paul et al., 2008).

2.2 Sectoral Pricing Dynamics

Electricity prices vary significantly by consuming sector—primarily residential, commercial, and industrial. Industrial consumers consistently face lower per-unit electricity costs compared to residential consumers. This disparity is generally attributed to economies of scale and load profiles; industrial facilities draw large, constant base loads, which are cheaper for utilities to supply compared to the highly variable, peak-heavy load profiles of residential neighborhoods (Faruqui et al., 2012). Furthermore, industrial customers often

receive power at higher transmission voltages, avoiding the distribution costs that are passed on to residential consumers (Borenstein, 2012).

2.3 Demographic and Infrastructural Influences

The structural characteristics of a utility's customer base also impact pricing. While a larger total customer base might suggest economies of scale, the density of those customers is a more critical metric. Research indicates that urban areas with high customer density often require complex, underground distribution networks and frequent maintenance, driving up infrastructural costs that are reflected in consumer pricing (Sioshansi, 2011).

By integrating these theoretical frameworks, the current study utilizes multiple linear regression to quantify the concurrent effects of region, sector, and density, providing updated empirical evidence for the U.S. electricity market.

3. Methodology

3.1 Data Sources and Variables

Data for this cross-sectional study were aggregated from two primary sources: the U.S. Energy Information Administration (EIA) and the U.S. Census Bureau. Monthly average state-level electricity pricing data were obtained from the EIA for January 2023 and January 2024 to allow for temporal examination. State area data (in square miles) were sourced from the U.S. Census Bureau to calculate customer density. Prior to analysis, the dataset underwent preprocessing to handle missing values and identify outliers.

The response variable is the **Average Price of Electricity** (measured in cents per kilowatt-hour [kWh]). The predictor variables included:

- **Sector** (Categorical: Residential, Industrial, Commercial)
- **Region** (Categorical: Northeast, South, Midwest, West)
- **Time** (Categorical: January 2023, January 2024)
- **Number of Ultimate Customers** (Numeric)
- **Customer Density** (Numeric: Number of customers per square mile)

3.2 Statistical Methods

Summary statistics (mean, standard deviation, minimum, maximum) and Karl-Pearson correlation coefficients were calculated to understand the baseline distribution and bivariate relationships. Multi-way Analysis of Variance (ANOVA) was utilized to test the significance of factor effects on the response variable.

To model the relationship between the predictors and electricity pricing, Multiple Linear Regression (MLR) was employed. Model adequacy was assessed using residual versus fitted plots, normal Q-Q plots, and standardized residual plots. Because the initial model violated assumptions of constant variance and normality of errors, a Box-Cox transformation was applied to the dependent variable to stabilize variance and achieve normality.

4. Results

4.1 Summary Statistics and Correlation

Table 1 presents the summary statistics for the average price of electricity across different sectors, regions, and time periods. The residential sector exhibited the highest average price (16.27 ¢/kWh), followed by the commercial (13.15 ¢/kWh) and industrial (9.95 ¢/kWh) sectors. Geographically, the Northeast reported the highest average prices (18.89 ¢/kWh), while the Midwest (10.63 ¢/kWh) and South (10.72 ¢/kWh) were the lowest. State-to-state variability was highest in the West region (SD = 9.46). Temporal comparisons showed negligible pricing increments between January 2023 (13.12 ¢/kWh) and January 2024 (13.13 ¢/kWh).

Table 1: Comparison of Average Electricity Price by Predictor Variables

Factor	Levels	Min	Max	Mean	SD
Sector	Commercial	7.80	43.65	13.15	5.78
	Industrial	5.47	40.19	9.95	5.57
	Residential	9.35	44.96	16.27	6.95
Region	Midwest	6.21	18.34	10.63	2.88
	Northeast	7.50	31.83	18.89	6.21
	South	5.75	17.48	10.72	2.73
	West	5.47	44.96	14.39	9.46
Time	January 2023	5.90	44.96	13.12	6.72
	January 2024	5.47	44.28	13.13	6.57
Overall		5.47	44.96	13.12	6.63

Correlation analysis revealed a weak positive correlation between the price of electricity and the number of customers ($r = 0.246$) and a moderate positive correlation with customer density ($r = 0.411$).

4.2 Initial Regression and Model Diagnostics

An initial regression model incorporating all predictors (including Time) explained 67.81% of the variability (Adjusted $R^2 = 0.6781$). ANOVA results indicated that "Time" and all its interaction terms were highly insignificant ($p = 0.992$); thus, Time was dropped from subsequent modeling.

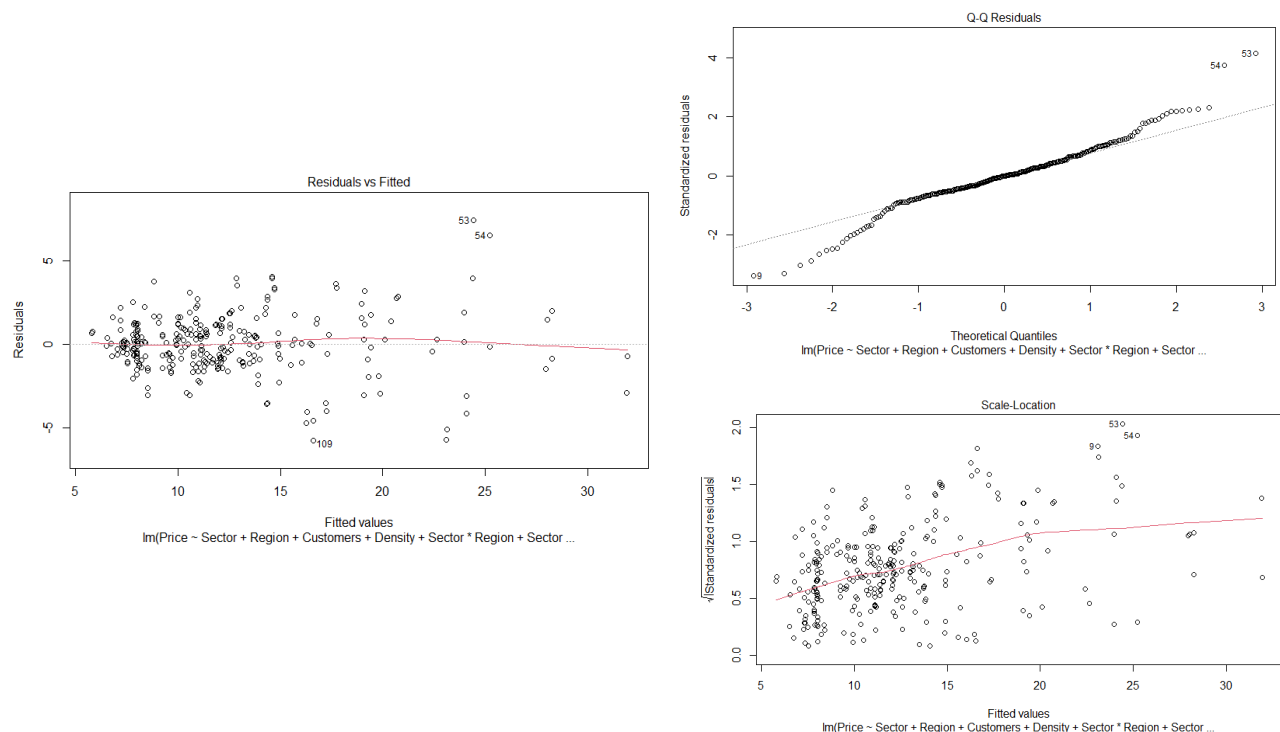


Figure 1: Diagnostic tests of the initial model

Diagnostic testing of the rebuilt model as shown in Figure 1 revealed a violation of the constant variance assumption (heteroscedasticity) and non-normality of errors. The distribution of the raw pricing data was positively skewed.

4.3 Data Transformation and Final Model

To address the violation of assumptions, potential outliers were removed, and a Box-Cox transformation was conducted. The log-likelihood curve shown in Figure 2 suggested an optimal λ between -0.3 and 0.2. A log transformation ($\lambda = 0$) was adopted for its interpretability.

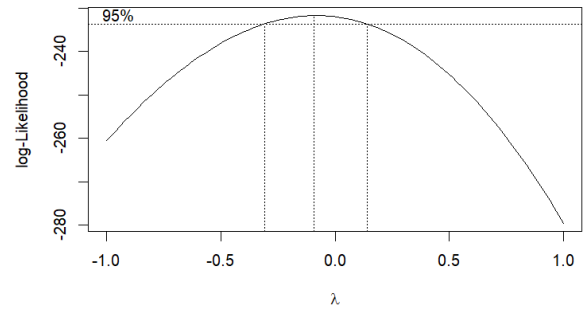


Figure 2: Box-Cox Transformation Plot

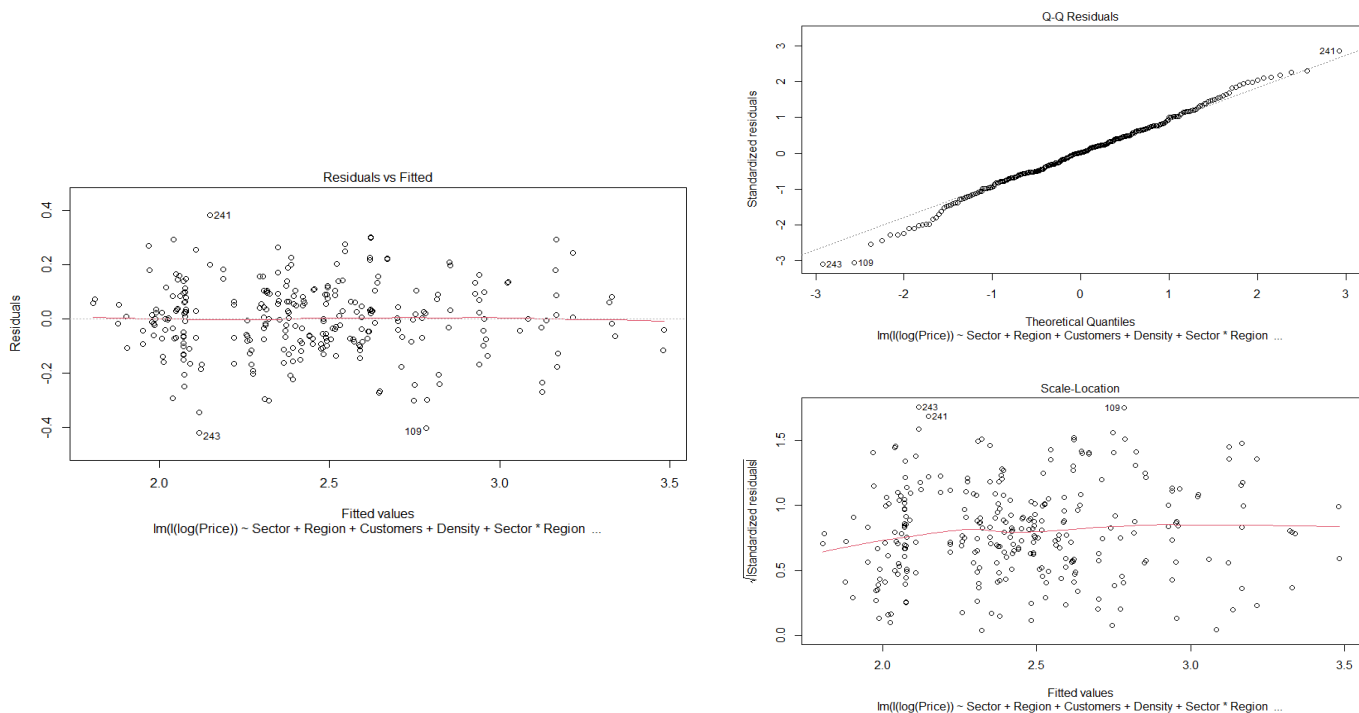


Figure 3: Diagnostic tests of the final model

The log-transformed model resolved the adequacy issues as shown in Figure 3, exhibiting constant variance among residuals. The final model yielded an Adjusted R^2 of 0.8536, indicating that Sector, Region, Number of Customers, Customer Density, and their interactions explain over 85% of the variance in U.S. electricity pricing.

After applying log transformation to the response variable, the regression model has been fitted. The ANOVA table for the log transformed model has been presented in Table 2.

Table 2: ANOVA for Log-Transformed Regression Model

Source	Df	Sum Sq	Mean Sq	F value	Pr(>F)
Sector	2	13.676	6.838	344.18	< 2.2e-16 ***
Region	3	12.746	4.248	213.85	< 2.2e-16 ***
Customers	1	0.851	0.851	42.86	3.35e-10 ***
Density	1	0.222	0.222	11.18	0.00095 ***
Sector:Region:Customers	6	2.628	0.438	22.04	< 2.2e-16 ***
Residuals	248	4.927	0.019		

The p-values for all the selected factors are found to be very small, the selected factors are highly significant for the model.

4.4 Regression Coefficients and Interpretation

Table 3 presents the regression coefficients for the final log-transformed model predicting the price of electricity. In this model, the categorical variables are assessed relative to a baseline reference group, which is the **Residential sector in the Northeast region**.

Table 3: Regression Coefficients for Log-Transformed Model

Predictor	Estimate	Std. Error	t value	Pr(> t)	Signif.
(Intercept)	3.114e+00	7.763e-02	40.119	< 2e-16	***
Sector: Commercial	-1.491e-01	9.945e-02	-1.499	0.135154	
Sector: Industrial	-5.264e-01	9.801e-02	-5.371	1.80e-07	***
Region: Midwest	-8.723e-01	1.067e-01	-8.174	1.54e-14	***
Region: South	-6.892e-01	9.819e-02	-7.019	2.13e-11	***
Region: West	-7.291e-01	1.055e-01	-6.911	4.05e-11	***
Customers	4.040e-08	2.792e-08	1.447	0.149158	
Density	1.446e-06	4.699e-07	3.077	0.002327	**
Sector(Comm):Region(Midwest)	8.285e-02	1.375e-01	0.602	0.547464	
Sector(Ind):Region(Midwest)	3.645e-01	1.295e-01	2.815	0.005264	**

Predictor	Estimate	Std. Error	t value	Pr(> t)	Signif.
Sector(Comm):Region(South)	7.520e-02	1.297e-01	0.580	0.562691	
Sector(Ind):Region(South)	4.608e-02	1.236e-01	0.373	0.709672	
Sector(Comm):Region(West)	1.281e-03	1.318e-01	0.010	0.992251	
Sector(Ind):Region(West)	2.531e-01	1.306e-01	1.939	0.053690	.
Sector(Comm):Customers	-2.807e-07	9.620e-08	-2.918	0.003851	**
Sector(Ind):Customers	-4.776e-05	7.136e-06	-6.693	1.45e-10	***
Region(Midwest):Customers	1.083e-07	4.955e-08	2.185	0.029804	*
Region(South):Customers	-2.710e-08	3.156e-08	-0.859	0.391227	
Region(West):Customers	2.538e-07	9.583e-08	2.648	0.008605	**
Customers:Density	-5.850e-13	1.757e-13	-3.330	0.001000	**
Sector(Comm):Density	-1.331e-06	1.894e-06	-0.703	0.482757	
Sector(Ind):Density	5.345e-04	7.759e-05	6.888	4.62e-11	***
Region(Midwest):Density	7.514e-07	3.193e-06	0.235	0.814140	
Region(South):Density	2.555e-07	7.829e-07	0.326	0.744431	
Region(West):Density	-1.839e-05	5.435e-06	-3.384	0.000831	***
Sector(Comm):Region(MW):Cust	7.666e-07	3.332e-07	2.301	0.022216	*
Sector(Ind):Region(MW):Cust	4.580e-05	8.607e-06	5.321	2.30e-07	***
Sector(Comm):Region(South):Cust	1.691e-07	1.161e-07	1.457	0.146458	
Sector(Ind):Region(South):Cust	4.625e-05	7.156e-06	6.464	5.38e-10	***
Sector(Comm):Region(West):Cust	7.619e-07	2.519e-07	3.024	0.002754	**
Sector(Ind):Region(West):Cust	6.500e-05	7.660e-06	8.485	1.98e-15	***
Sector(Comm):Region(MW):Density	-6.824e-06	1.520e-05	-0.449	0.653860	
Sector(Ind):Region(MW):Density	-4.773e-04	3.406e-04	-1.401	0.162352	

Predictor	Estimate	Std. Error	t value	Pr(> t)	Signif.
Sector(Comm):Region(South):Den	9.754e-06	4.586e-06	2.127	0.034427	*
Sector(Ind):Region(South):Den	-1.851e-04	1.614e-04	-1.146	0.252775	
Sector(Comm):Region(West):Den	-2.419e-05	3.576e-05	-0.676	0.499462	
Sector(Ind):Region(West):Den	-2.579e-03	4.604e-04	-5.601	5.66e-08	***
Region(MW):Customers:Density	-3.636e-13	6.359e-13	-0.572	0.567967	
Region(South):Customers:Density	5.064e-13	2.262e-13	2.239	0.026069	*
Region(West):Customers:Density	-8.501e-13	7.699e-13	-1.104	0.270586	

Main Effects Interpretation: The (Intercept) value of 3.114 represents the average log-price for the baseline group (Residential, Northeast). Examining the main effects, the industrial sector has a highly significant negative coefficient (-0.526, $p < 0.001$), indicating that the average price of electricity for industrial use is substantially lower than for residential use, holding other variables constant. Geographically, compared to the Northeast, all other regions—the Midwest (-0.872), South (-0.689), and West (-0.729)—exhibit significantly lower average electricity prices ($p < 0.001$).

While the absolute number of customers alone does not show a statistically significant main effect ($p = 0.149$), customer density is highly significant and positive (1.446×10^{-6} , $p = 0.002$). This suggests that an increase in customer density generally drives up the average price of electricity, a finding consistent with the high capital costs of dense urban infrastructure networks.

Two-Way Interaction Effects: The model reveals several important nuances through two-way interactions:

- **Sector and Region:** The interaction term SectorIndustrial:RegionMidwest is significant and positive (0.364, $p = 0.005$). This means that while industrial rates are lower overall and Midwest rates are lower overall, the specific combination of industrial pricing within the Midwest does not drop *as sharply* as the independent main effects would predict—partially offsetting the steep discounts.
- **Sector and Customers:** The interaction between the industrial sector and the number of customers (SectorIndustrial:Customers) is significantly negative (-4.776×10^{-5} , $p < 0.001$). This implies that as the customer base grows, the price discount realized by the industrial sector compared to the residential sector deepens.

- **Sector and Density:** Higher density increases prices overall, but the SectorIndustrial:Density interaction indicates that this penalty is drastically amplified for the industrial sector (5.345×10^{-4} , $p < 0.001$) relative to the residential baseline.
- **Region and Density:** Conversely, the RegionWest:Density interaction is highly negative (-1.839×10^{-5} , $p < 0.001$), showing that the cost-inflating effect of density seen in the Northeast baseline is mitigated (or reversed) in the West.

Three-Way Interaction Effects: Complex three-way interactions indicate fine-grained state-level dynamics. For example, the effect of an increasing customer base on industrial pricing varies significantly depending on the region (e.g., SectorIndustrial:RegionWest:Customers is highly significant and positive at $p < 0.001$). Similarly, the impact of density on industrial rates in the West (SectorIndustrial:RegionWest:Density) exhibits a strong negative offset (-2.579×10^{-3} , $p < 0.001$). These multi-level dependencies confirm that electricity pricing cannot be generalized purely by sector or region alone; the structural demographics of the customer base dictate localized variations.

5. Discussion

The statistical analysis confirms several economic theories regarding U.S. electricity markets. First, the finding that commercial and industrial sectors have significantly cheaper electricity than the residential sector aligns with established literature on load profiling (Faruqui et al., 2012). Industrial clients provide stable base loads and take power at high voltages, reducing distribution costs for the utility.

Second, geographical location is a primary driver of pricing. The Northeast's significantly higher prices reflect the region's historical infrastructural constraints, such as pipeline bottlenecks for natural gas, and higher overall costs of living and regulatory compliance (Joskow, 2006). In contrast, the South and Midwest benefit from proximity to cheaper traditional and renewable generation sources.

Third, the analysis highlighted the significant positive impact of customer density on pricing. While it is often assumed that dense urban areas benefit from economies of scale, the positive correlation suggests that the high costs of installing, upgrading, and maintaining complex urban grid infrastructure (e.g., underground cabling) outweigh the efficiency of shorter transmission distances (Sioshansi, 2011).

Finally, the lack of significant temporal variation between January 2023 and January 2024 suggests short-term stability in the macroeconomic factors governing electricity generation costs during this period, despite broader inflationary pressures in the economy.

6. Conclusion and Policy Implications

This study demonstrates that electricity pricing in the USA is a multifaceted issue driven primarily by sector classification, geographic region, and customer density. By employing a robust log-transformed multiple linear regression model, the study successfully explained 85.36% of the variance in state-level pricing.

6.1 Recommendations and Policy Implications

1. **Sector-Specific Rate Design:** Utilities and regulators must continue to refine rate structures based on sector types, recognizing that residential consumers bear a disproportionate share of grid distribution costs. Policies promoting residential demand-response programs could help smooth load profiles and lower costs.
2. **Targeted Regional Infrastructure Investment:** The high costs in the Northeast underscore the need for infrastructure investments (e.g., transmission lines connecting to cheaper generation hubs) to alleviate regional price disparities.
3. **Density Management:** Given that increased customer density correlates with higher prices, urban planners and utilities should explore smart-grid technologies that optimize urban distribution networks to mitigate the high maintenance costs associated with dense service areas.

6.2 Limitations and Future Research

A limitation of this study is the potential for model overfitting, as the dataset was not split into training and testing subsets for validation. Future research should employ cross-validation techniques to build more robust predictive models. Additionally, incorporating variables such as state-level renewable energy penetration, real-time fuel costs, and deregulation status could further enhance the explanatory power of the model.

References

- Borenstein, S. (2012). The redistributive impact of nonlinear electricity pricing. *American Economic Journal: Economic Policy*, 4(3), 56-90.
- Borenstein, S., & Bushnell, J. (2015). The US electricity industry after 20 years of restructuring. *Annual Review of Economics*, 7(1), 437-463.
- Faruqui, A., Hledik, R., & Palmer, J. (2012). Time-varying and dynamic rate design. *Global Power Best Practice Series*. The Regulatory Assistance Project (RAP).
- Joskow, P. L. (2006). Markets for power in the United States: An interim assessment. *The Energy Journal*, 27(1), 1-36.

Paul, A., Burtraw, D., & Palmer, K. (2008). Compensation for electricity consumers under a U.S. CO₂ emissions cap. *Resources for the Future Discussion Paper*, DP-08-25.

Sioshansi, F. P. (Ed.). (2011). *Smart Grid: Integrating Renewable, Distributed & Efficient Energy*. Academic Press.